IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

James L. Hedrick
11/19/24

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

Summary of methodologies

Data Collection:

- Extracted SpaceX launch data and cleaned/standardized it for analysis.

EDA:

- Visualized launch outcomes, success rates, and payload patterns using SQL and Python libraries.

Interactive Tools:

- Built a Folium map to analyze launch site proximities.
- Created a Plotly Dash dashboard for real-time success/payload correlations.

Predictive Modeling:

- Trained Logistic Regression, SVM, Decision Tree, and KNN models.
- Optimized hyperparameters using GridSearchCV and evaluated performance.

Summary of all results

EDA:

- **KSC LC-39A** has the highest successful launches.
- Payloads between **2000–4000 kg** yield the best success rates.
- F9 FT boosters outperform others.

Interactive Analysis:

- Geographic location minimally impacts success rates.
- Booster version and payload mass are critical factors.

Machine Learning:

- **Decision Tree** achieved the highest accuracy (**86.07%**).
- Other models performed comparably but with slightly lower accuracy.

Introduction

Project background and context

The commercial space industry is rapidly evolving, with companies like SpaceX revolutionizing the cost and efficiency of space travel. One of SpaceX's key innovations is the reusability of its Falcon 9 first stage, significantly reducing launch costs compared to competitors. This project focuses on analyzing SpaceX's launch data to extract meaningful insights and predict the likelihood of reusing the first stage. These insights aim to support the hypothetical Space Y in competing effectively in the commercial space industry.

Problems you want to find answers

- **Launch Success Factors**: What factors contribute most to a successful landing of the Falcon 9 first stage?
- **Site and Payload Patterns**: Which launch sites and payload ranges are associated with the highest success rates?
- **Predictive Modeling**: Can machine learning models reliably predict the success of future SpaceX launches?

Section 1

Methodology

Methodology

Executive Summary

Data Collection Methodology

Data Sources:

- launch data was collected from publicly available repositories, including SpaceX's website and associated datasets.

Tools Used:

- Python library (pandas) was used to load and manipulate the data.

Data Wrangling

Processing:

- Cleaned missing and inconsistent data.
- Standardized column formats for easy analysis.

Preparation:

- Converted necessary columns into usable formats.

Exploratory Data Analysis (EDA)

Visualization:

- Used matplotlib and seaborn for initial visual insights.
- Focused on correlations between variables like launch site, payload, and mission outcomes.

SQL Analysis:

- Performed queries on cleaned datasets to gain deeper insights into trends (e.g., launch success rates by site).

Methodology

Executive Summary (Continued)

Interactive Visual Analytics

Folium:

- Mapped launch site proximities to geographical features (e.g., highways, coastlines).

Plotly Dash:

- Created interactive dashboards with dropdowns and sliders for real-time data exploration.

Predictive Analysis

Classification Models:

- Models: Logistic Regression, Support Vector Machines (SVM), Decision Trees, and K-Nearest Neighbors.
- Grid Search and Cross-Validation: Tuned hyperparameters and optimized model performance.

Evaluation:

- Accuracy scores and confusion matrices were used to evaluate model performance.
- Identified the best-performing model to predict first-stage reuse likelihood.

Data Collection

Data Sources:

- **SpaceX REST API**: Collected launch data including rockets, payloads, and landing outcomes via API endpoints (/v4/launches/past).
- **Web Scraping**: Extracted Falcon 9 launch data from Wiki pages using the BeautifulSoup library to parse HTML tables.

Data Collection and Preparation

- **API Integration**: Used Python's requests library to perform GET requests to retrieve JSON responses.
  - Normalized JSON data into a tabular format using json_normalize.
- **Web Scraping**: Parsed HTML tables from Wikipedia to extract launch data.
  - Converted parsed data into Pandas DataFrames for further analysis.
- **Filtering and Cleaning**: Filtered out Falcon 1 launches to focus solely on Falcon 9.
  - Handled NULL values:
    - Replaced missing values in PayloadMass with the column's mean.
    - Retained NULLs in LandingPad for further processing using one-hot encoding.

Data Collection – SpaceX API

API Source:

- SpaceX REST API (api.spacexdata.com/v4/launches/past) was utilized to retrieve Falcon 9 launch data.

Data Format:

- JSON response with detailed information about launches, rockets, payloads, and landing outcomes.

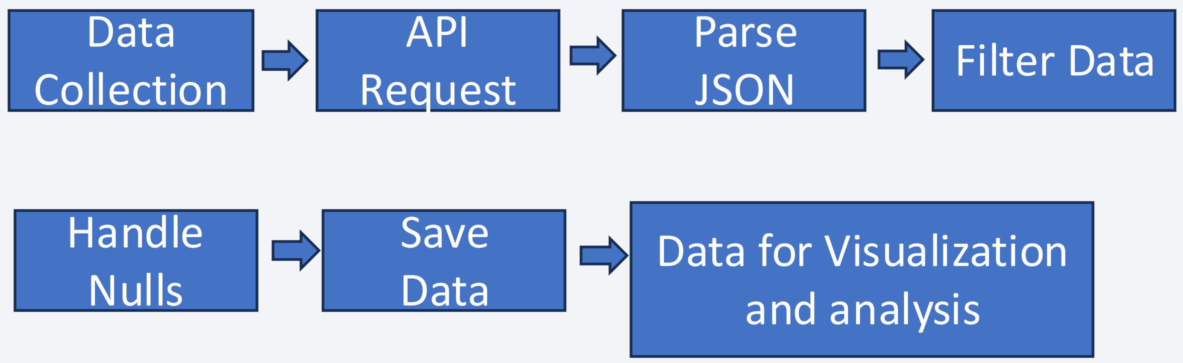
Retrieval Process:

- Sent GET request using
- Python's requests library. Parsed the JSON response using pandas.json_normalize to convert structured JSON into a flat table.

Data Cleaning:

- Removed Falcon 1 launches using filters.
- Dealt with null values in columns like PayloadMass by replacing them with the mean.
- Left columns like LandingPad with nulls for one-hot encoding.

Flowchart of SpaceX API



- https://github.com/JLHedrick/Data-Science-Capstone-Project/blob/main/jupyter-labs-spacex-data-collection-api.ipynb

Data Collection - Scraping

Data Source: Wikipedia page containing Falcon 9 launch records in HTML tables.

Tools Used: Python's requests library for fetching HTML content and BeautifulSoup for parsing the HTML.

Web Scraping Process: Sent GET request to the Wikipedia page URL.

- Used BeautifulSoup to parse and locate HTML <table> tags containing the required launch data.

Data Extraction: Extracted table rows (<tr>) and individual cells (<td>) using loops.

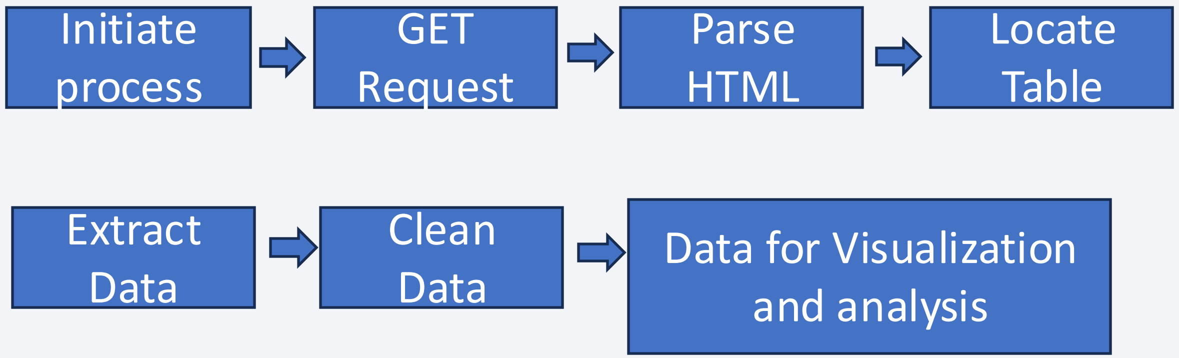
- Collected columns like launch date, site, payload, and landing outcomes.

Data Cleaning:Parsed string data for numerical conversions (e.g., extracting numbers from text fields).

- Removed irrelevant or incomplete rows.

Output: Stored in a pandas dataframe for further analysis and visualization.

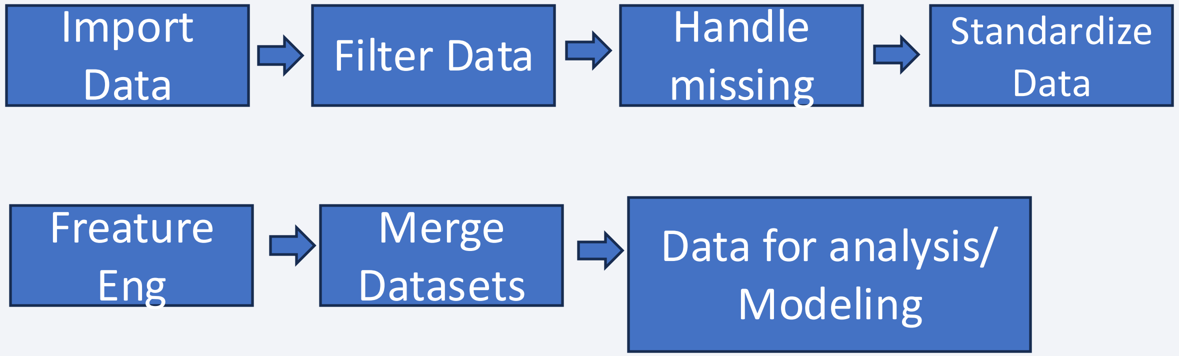
Flowchart of web scraping



https://github.com/JLHedrick/Data-Science-Capstone-Project/blob/main/jupyter-labs-webscraping.ipynb

Data Wrangling

- **Data Cleaning**: Removed irrelevant columns and filtered data to include only Falcon 9 launches.
- Handled missing values, replacing nulls in Payload Mass with column mean.
- Dropped unnecessary rows related to Falcon 1.
- **Data Transformation**: Standardized numerical columns (e.g., Payload Mass) using StandardScaler.
- Created new columns using one-hot encoding for categorical data like LandingPad.
- Extracted and merged data from multiple APIs for attributes like boosters and payloads.
- **Data Integration**: Combined data from SpaceX REST API and web-scraped datasets into a single pandas dataframe.
- Ensured consistency in column types and data formats.
- **Output**: Saved the processed dataframe for EDA, visualization, and predictive analysis.



https://github.com/JLHedrick/Data-Science-Capstone-Project/blob/main/labs-jupyter-spacex-Data%20wrangling-2.ipynb

EDA with Data Visualization

Pie Chart:

To visualize the distribution of successful and failed launches across all SpaceX launch sites.

Purpose: Help identify which sites had the highest success rates.

Scatter Plot:

To examine the relationship between payload mass and launch success, with data points colored by booster version category.

Purpose: To determine if payload mass impacts the likelihood of a successful launch and to observe performance differences between boosters.

Line Chart:

(If created) To analyze time-series data for launches, tracking success rates over time.

Purpose: To spot trends or improvements in SpaceX's technology and processes over time.

Bar Chart:

To compare the number of launches per launch site.

Purpose: To identify the most frequently used launch sites and their operational activity.

Folium Map (Interactive):

To display geographical locations of launch sites with markers for success or failure outcomes.

Purpose: To explore geographical factors influencing launch outcomes, such as proximity to coasts.

https://github.com/JLHedrick/Data-Science-Capstone-Project/blob/main/edadataviz.ipynb

EDA with SQL

- Names of launch sites: %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
- Launch begins with CAA: %sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
- Total payload mass: %sql SELECT SUM("PAYLOAD_MASS__KG_") AS total_payload_mass FROM SPACEXTABLE WHERE "Customer" LIKE '%NASA (CRS)%';
- Average payload by booster F9 v1.1: %sql SELECT AVG("PAYLOAD_MASS__KG_") AS average_payload_mass FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
- Date of successful landing: %sql SELECT MIN("Date") AS first_successful_ground_pad_landing_date FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)';
- Names of boosters 4000 but less than 6000: %sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
- Total number of successful/failures: %sql SELECT CASE WHEN "Mission_Outcome" LIKE 'Success%' THEN 'Success' WHEN "Mission_Outcome" LIKE 'Failure%' THEN 'Failure' ELSE "Mission_Outcome" END AS outcome_category, COUNT(*) AS outcome_count FROM SPACEXTABLE GROUP BY outcome_category;
- Boosters with max payloads: %sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);
- list records of month names 2015: %sql SELECT substr("Date", 6, 2) AS month, "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Failure (drone ship)' AND substr("Date", 1, 4) = '2015';
- Rank the landing outcomes: %sql SELECT "Landing_Outcome", COUNT(*) AS outcome_count FROM SPACEXTABLE WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY outcome_count DESC;

https://github.com/JLHedrick/Data-Science-Capstone-Project/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Build an Interactive Map with Folium

Markers:

- Markers were added for each launch site.
- These markers represent the locations of the SpaceX launch sites. They allow users to visualize the geographical distribution of the sites.

Circles:

- Circles were placed around each launch site.
- The circles provide a spatial context, such as the proximity to landmarks like highways, railways, or coastlines, helping in visualizing the surrounding area of the launch sites.

Popup Information:

- Popups were added to the markers to display detailed information, such as the name of the launch site and additional statistics like the success rate.
- These popups give users contextual details when interacting with the map.

Lines:

- Lines connecting launch sites to specific points of interest, such as nearby cities, highways, or coastlines.
- These lines demonstrate proximity and accessibility, which could help in logistical planning or understanding the relationship between launch site location and infrastructure.

Mouse Position Plugin:

- A plugin that displays the coordinates of the mouse pointer.
- This feature aids in exploring the map and identifying the geographical coordinates of points of interest.

https://github.com/JLHedrick/Data-Science-Capstone-Project/blob/main/lab_jupyter_launch_site_location.ipynb

Build a Dashboard with Plotly Dash

https://github.com/JLHedrick/Data-Science-Capstone-Project/blob/main/Capstone_dashboard.py

Success Pie Chart:

- A pie chart displaying the success rates of launches per launch site or overall.
- To give users a quick visual representation of the proportion of successful versus failed launches across sites or for a specific site. This helps in identifying performance trends and reliability.

Success-Payload Scatter Plot:

- A scatter plot showing the correlation between payload mass and launch success.
- To analyze the relationship between payload mass and the likelihood of success, helping to identify any patterns or limits in payload capacity for successful launches.

Interactive Dropdown Menu:

- A dropdown allowing users to select specific launch sites or view data for all sites.
- To enable user-driven exploration of data, providing flexibility to focus on a specific launch site or to observe overall trends.

Payload Range Slider:

- A slider to filter data based on payload mass range.
- To allow users to analyze success rates for specific payload ranges, offering insights into how payload mass affects success outcomes.

Predictive Analysis (Classification)

Building Models:

- Utilized **Logistic Regression**, **Support Vector Machines (SVM)**, **Decision Trees**, and **K-Nearest Neighbors (KNN)** to classify the likelihood of a successful rocket landing.
- Created **parameter grids** specific to each model for hyperparameter tuning.

Data Preparation:

- Standardized the dataset using StandardScaler to ensure all features were on the same scale.
- Split the data into **training** (80%) and **test** (20%) sets using train_test_split.

Hyperparameter Tuning:

- Employed **GridSearchCV** with **cross-validation (cv=10)** to find the optimal parameters for each classification model.
- Evaluated models based on **best parameters** and cross-validation scores.

Model Evaluation:

- Measured the performance of each model using the **accuracy** on test data.
- Generated **confusion matrices** to visualize the classification results (True Positive, True Negative, etc.).
- Determined the best-performing model based on test accuracy.

Comparison and Selection:

- Compared the accuracy scores of Logistic Regression, SVM, Decision Trees, and KNN.
- Selected the **Decision Tree Classifier** as the best-performing model based on its high accuracy and well-defined parameters.

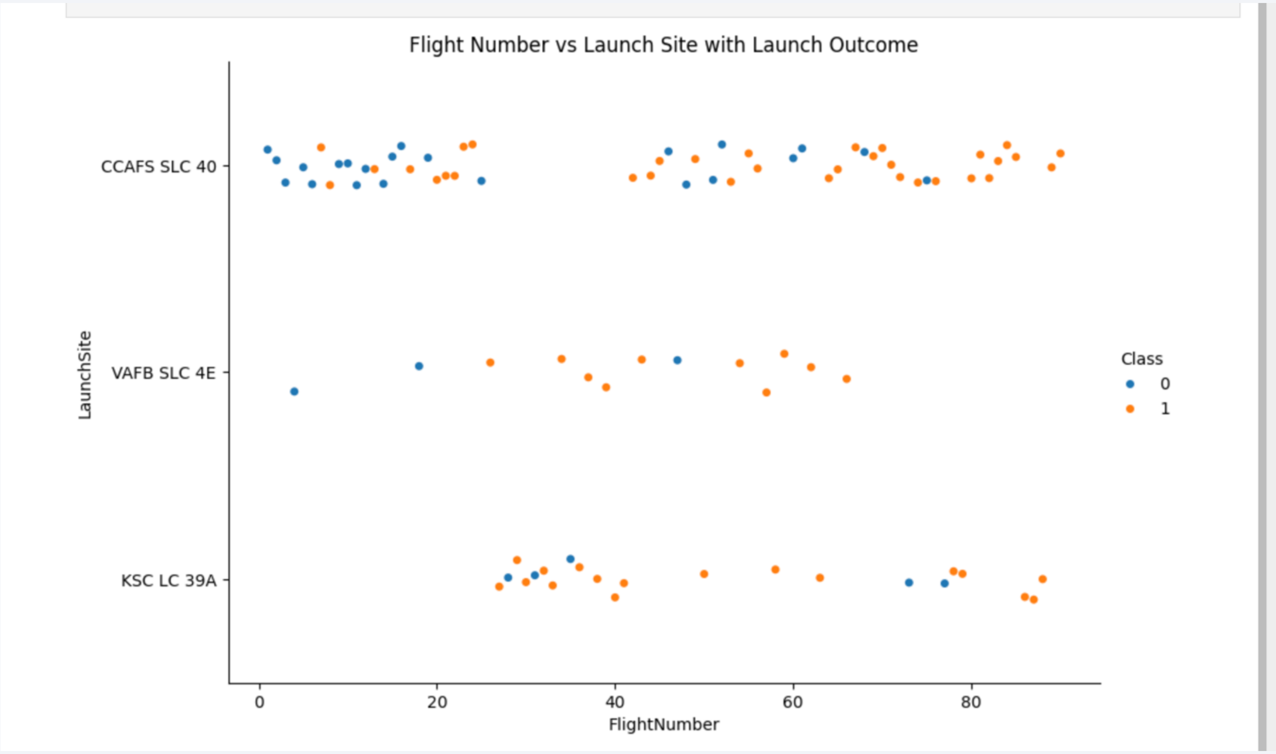
Results

- Exploratory data analysis results
- Interactive analytics demo in screenshots
- Predictive analysis results

Section 2

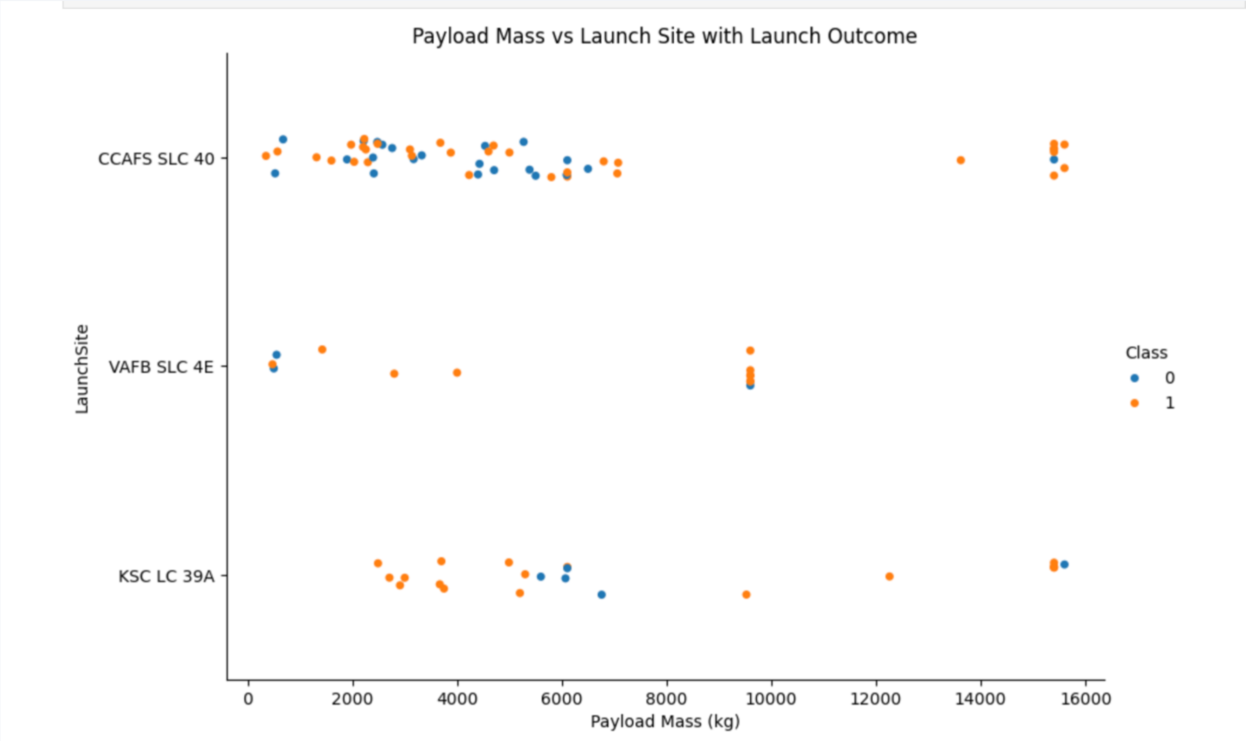
Insights drawn from EDA

Flight Number vs. Launch Site



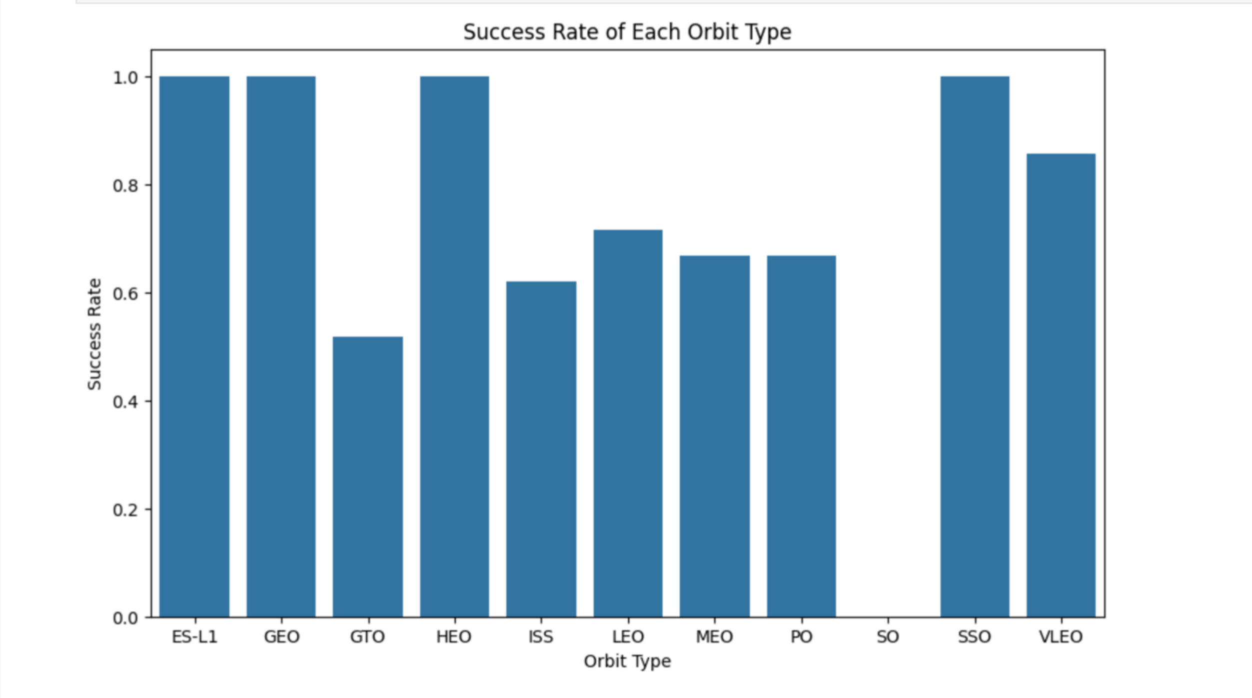
- The scatter plot of **Flight Number vs. Launch Site** is designed to visualize the frequency and sequence of launches at each launch site.
- The **Flight Number** represents the chronological order of launches. Higher flight numbers indicate more recent launches.
- Launch site, CCAFS SLC-40, has more frequent flight numbers that suggest it is the primary or preferred site for launches, while site, has limited number of recent flights.
- It provides an overview of the utilization of SpaceX's infrastructure across different sites.

Payload vs. Launch Site



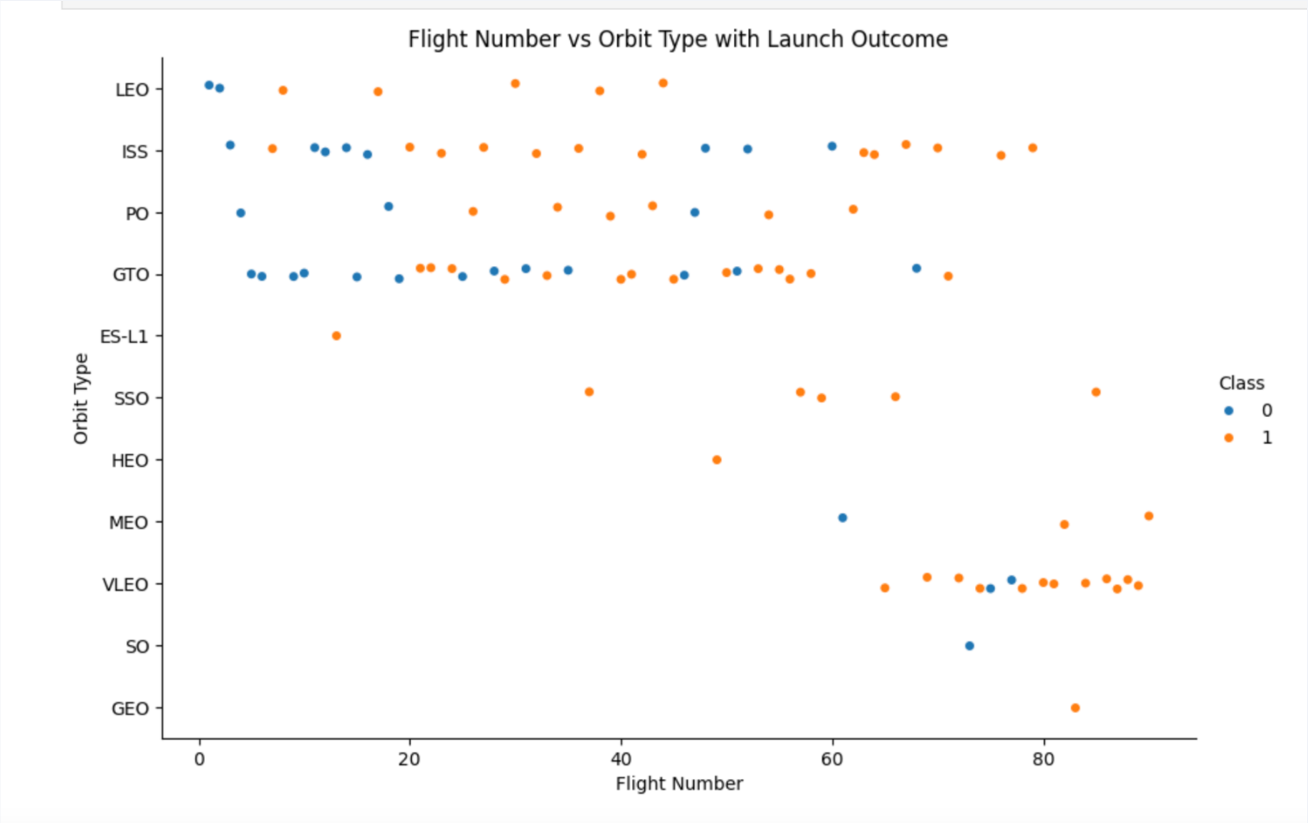
- This scatter plot visualizes the relationship between the **payload mass** (on the x-axis) and the **launch site** (on the y-axis), where It helps to identify which launch sites are used for missions with different payload capacities.
- Sites used for heavier payloads can often be seen with points farther to the right on the x-axis, and CCAFS SLC 40 has 1) the most launches and the highest concentration of high payloads
- KSC LC 39a has a constant and consistent across multiple payload launches.

Success Rate vs. Orbit Type



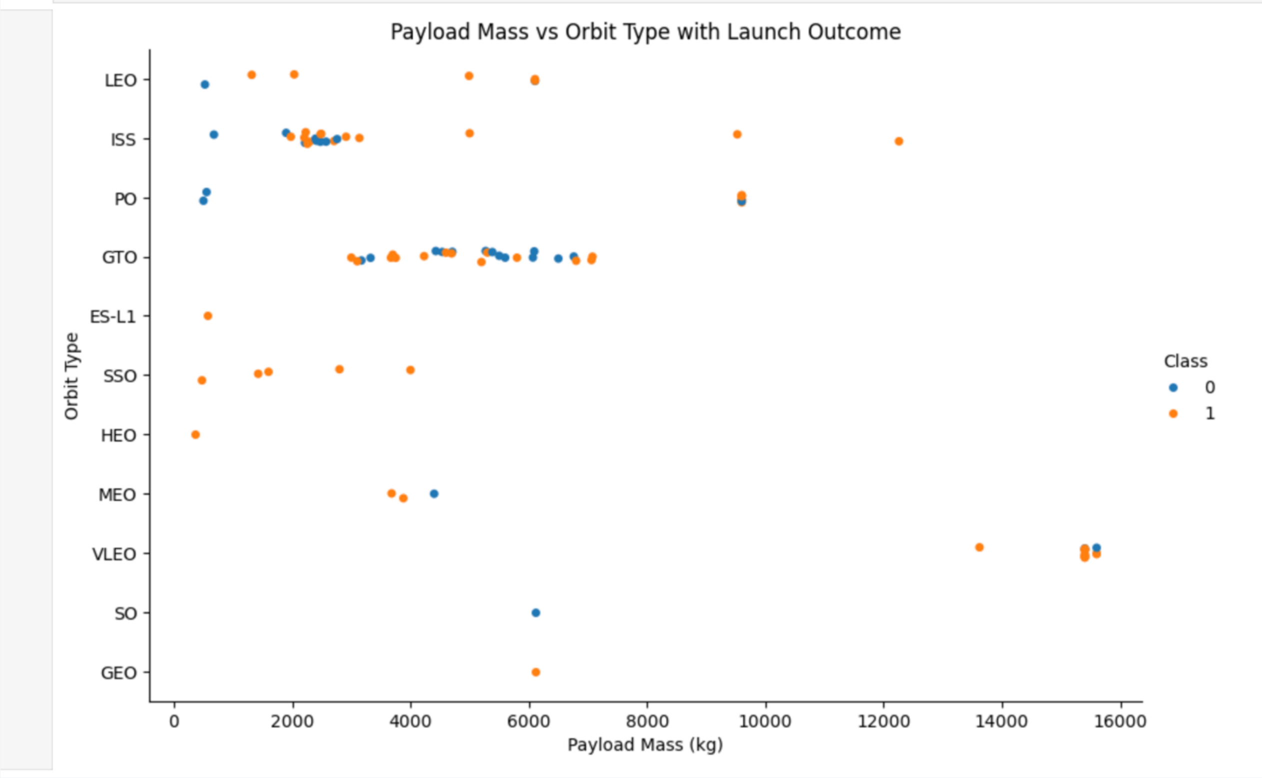
- The bar chart demonstrates the **success rate** of SpaceX launches for each orbit type.
- Each bar corresponds to a specific orbit type, with the height representing the success rate as a percentage.
- ES L1, GEO, HEO and SSO have the highest rates of success
- Higher success rates for specific orbits suggest that SpaceX has optimized operations for those mission profiles.
- Lower success rates may highlight areas for operational improvements or additional testing.

Flight Number vs. Orbit Type



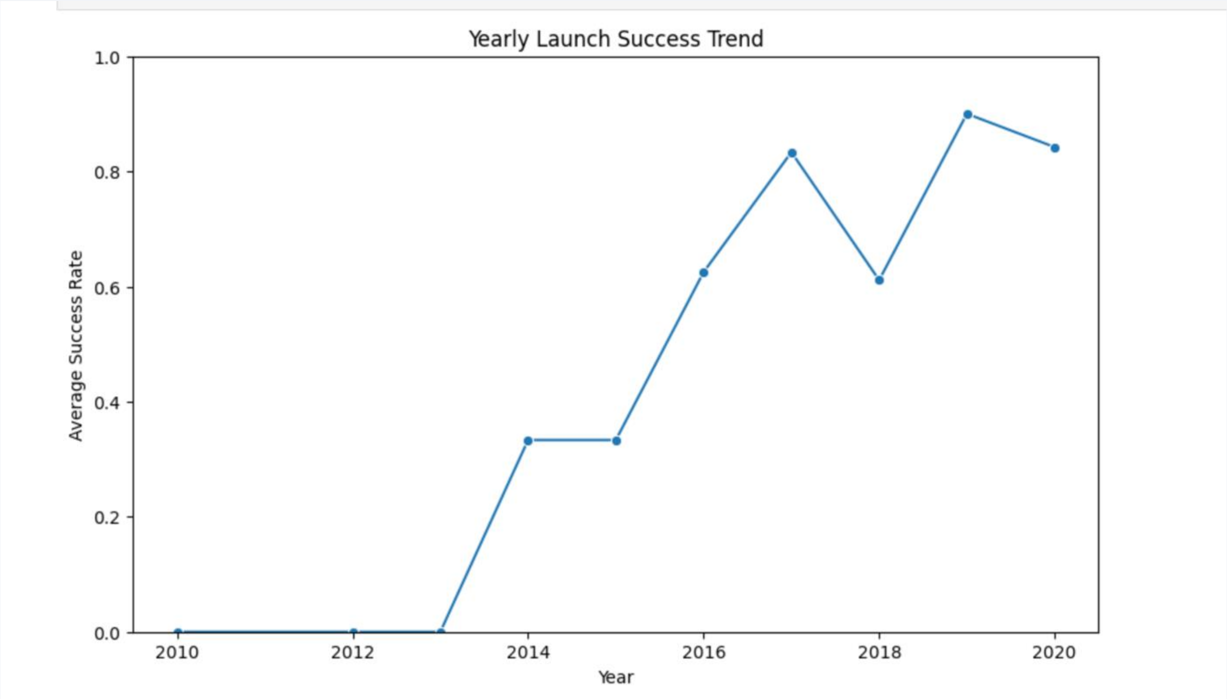
- Flight numbers increase over time, representing chronological mission history.
- Early flight numbers may concentrate around specific orbits (e.g., LEO, ISS ,PO and GTO) as SpaceX gained experience.
- Later flights SSO, VLEO and MEO contributed significantly
- Highlights focus areas over time, aiding future mission planning and strategy.

Payload vs. Orbit Type



- This scatter plot examines the relationship between payload mass (measured in kilograms) and the type of orbit the payload was destined for. It provides insights into how SpaceX's payload capabilities align with the requirements of various orbit types.
- GTO has a heavy concentration of medium payloads, while SSO, LEO and ISS had a higher concentration of lighter loads.
- Heavier payloads to higher orbits like GTO reflect more complex, high-energy missions requiring advanced engineering and larger booster stages

Launch Success Yearly Trend



- The purpose of this plot is to observe SpaceX's annual launch performance trends, focusing on the success rate of their missions over the years. It provides insights into the company's growth, technical advancements, and reliability improvements.
- **Early Years**: In the initial years, the success rate may have been inconsistent due to development phases and technology testing.
- **Later Years**: A noticeable upward trend in success rates may appear as SpaceX gained more experience, improved technology, and refined operations.
- Significant spikes in success rates may align with major milestones like the introduction of *Falcon 9* Block 5 (known for its reusability and reliability).

All Launch Site Names

- %sql SELECT DISTINCT "Launch_Site" FROM SPACEXTABLE;
- **Objective**: The query ensures no duplicates are included in the output, providing a concise list of all unique SpaceX launch sites.
- **Importance:** Knowing the unique launch sites helps analyze SpaceX's operational capabilities and geographical distribution of launches.

Launch Site Names Begin with 'CCA'

- %sql SELECT * FROM SPACEXTABLE WHERE "Launch_Site" LIKE 'CCA%' LIMIT 5;
- This query is used to retrieve the first 5 rows from the SPACEXTABLE where the Launch_Site column contains values starting with 'CCA'.
- **Analyze Specific Sites**: You can focus on data for launch sites starting with "CCA" (e.g., "CCAFS SLC-40").

Total Payload Mass

- %sql SELECT SUM("PAYLOAD_MASS__KG_") AS total_payload_mass FROM SPACEXTABLE WHERE "Customer" LIKE '%NASA (CRS)%';
- This query calculates the total payload mass (in kilograms) for launches where the customer includes **'NASA (CRS)'**.
- **Focus on NASA (CRS) Missions**: The query narrows down data to only those missions associated with NASA's Commercial Resupply Services (CRS).
- **Summarize Payloads**: The total payload mass is a critical metric for analyzing the scale of NASA (CRS) missions and evaluating their overall logistics.

Average Payload Mass by F9 v1.1

- %sql SELECT AVG("PAYLOAD_MASS__KG_") AS average_payload_mass FROM SPACEXTABLE WHERE "Booster_Version" = 'F9 v1.1';
- This query calculates the **average payload mass** (in kilograms) for all launches that used the **'F9 v1.1'** booster version.
- **Focus on Specific Booster:** This query isolates data for the **F9 v1.1 booster version** to analyze its performance or usage characteristics.
- **Understand Payload Trends**: Calculating the average payload mass helps evaluate the typical mission scale for a specific booster version, which could inform design, logistics, and future mission planning.

First Successful Ground Landing Date

- %sql SELECT MIN("Date") AS first_successful_ground_pad_landing_date FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (ground pad)’;
- This query retrieves the **earliest date** when a successful ground pad landing occurred, based on the data in the SPACEXTABLE.
- **Historical Insight**: This query identifies the starting point for successful ground pad landings, showcasing SpaceX's progress and milestones in reusability.
- **Landing Achievement Analysis**: Knowing when ground pad landings first succeeded helps assess SpaceX's technical growth and operational improvements.

Successful Drone Ship Landing with Payload between 4000 and 6000

- %sql SELECT DISTINCT "Booster_Version" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Success (drone ship)' AND "PAYLOAD_MASS__KG_" > 4000 AND "PAYLOAD_MASS__KG_" < 6000;
- This query retrieves all the **unique booster versions** that successfully landed on a drone ship, while carrying payloads weighing between **4000 kg and 6000 kg**.
- **Performance Analysis**: Helps identify which booster versions are capable of successfully landing on a drone ship with specific payloads, showcasing their reliability and reusability
- **Payload Capacity Insights**: Provides insights into the payload range handled by specific booster versions, aiding in future mission planning and engineering optimizations.

Total Number of Successful and Failure Mission Outcomes

- %sql SELECT CASE WHEN "Mission_Outcome" LIKE 'Success%' THEN 'Success' WHEN "Mission_Outcome" LIKE 'Failure%' THEN 'Failure' ELSE "Mission_Outcome" END AS outcome_category, COUNT(*) AS outcome_count FROM SPACEXTABLE GROUP BY outcome_category;
- This query categorizes mission outcomes into simplified **"Success"** and **"Failure"** groups (based on patterns in the Mission_Outcome column) and counts how many times each outcome occurs. Other outcomes that do not fit these patterns are kept as-is.
- **Simplifies Analysis**: Reduces complex mission outcome descriptions into meaningful categories (e.g., "Success" and "Failure").
- **Insight into Mission Trends**: Helps evaluate the overall mission success rate and failure patterns.

Boosters Carried Maximum Payload

- %sql SELECT "Booster_Version" FROM SPACEXTABLE WHERE "PAYLOAD_MASS__KG_" = (SELECT MAX("PAYLOAD_MASS__KG_") FROM SPACEXTABLE);
- This query identifies the **Booster Version** that carried the heaviest payload by selecting the row where the payload mass equals the maximum payload mass in the table.
- **Identify Top Performers**: Highlights the booster that managed the heaviest payload, which may indicate superior capabilities or technical advancements.
- **Evaluate Payload Capacity**: Useful for assessing the upper payload limits for specific booster versions.

2015 Launch Records

- %sql SELECT substr("Date", 6, 2) AS month, "Landing_Outcome", "Booster_Version", "Launch_Site" FROM SPACEXTABLE WHERE "Landing_Outcome" = 'Failure (drone ship)' AND substr("Date", 1, 4) = '2015';
- This query retrieves details of failed drone ship landings during the year 2015. Specifically: **Month** of the launch, the **Landing Outcome** (filtered to failures), the **Booster Version** use and the **Launch Site**.
- **Failure Insights**: Understanding which boosters and launch sites were involved in failed drone ship landings can guide future improvements in landing success.
- **Year-Specific Filtering**: Focusing on 2015 narrows the analysis to a specific period, which is useful for trend analysis.

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

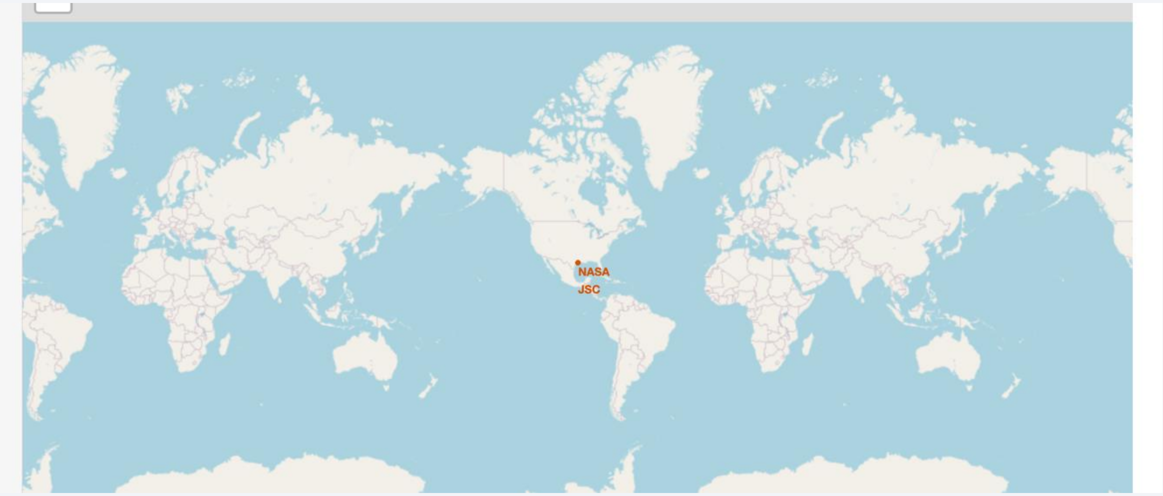
- %sql SELECT "Landing_Outcome", COUNT(*) AS outcome_count FROM SPACEXTABLE WHERE "Date" BETWEEN '2010-06-04' AND '2017-03-20' GROUP BY "Landing_Outcome" ORDER BY outcome_count DESC;
- This query retrieves the count of different landing outcomes for SpaceX launches that occurred between June 4, 2010, and March 20, 2017, grouped by the specific outcome type and sorted in descending order of their frequency.
- **Performance Analysis**: This query helps in understanding how frequently each type of landing outcome occurred during a specific time period.
- **Prioritization**: Identifying the most and least frequent outcomes can guide prioritization of operational improvements.

Section 3

Launch Sites Proximities Analysis

SpaceX Launch Sites on a Global Map

NASA



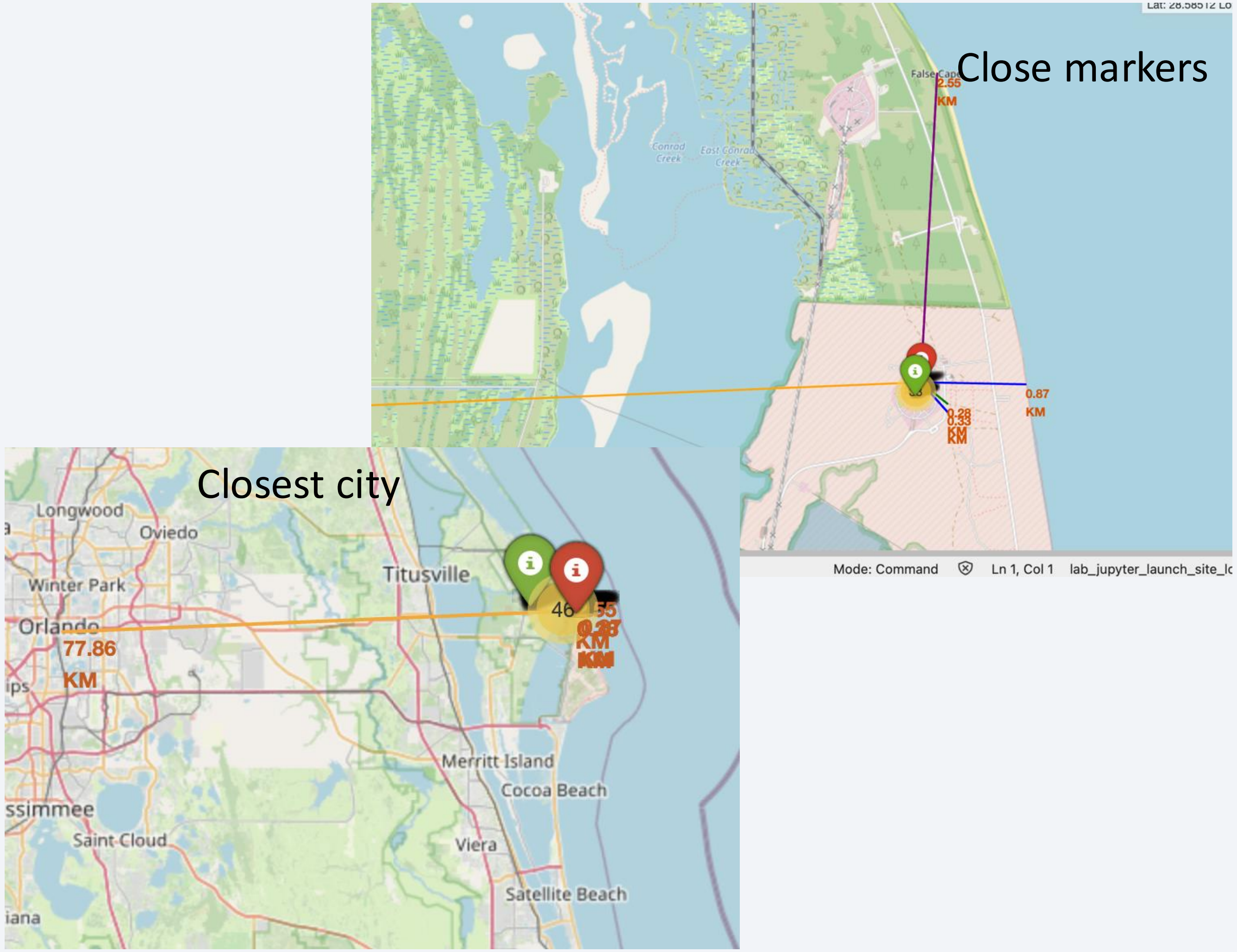
Launch sites

- SpaceX Launch Sites on a Global Map
- The purpose of this visualization is to show the geographical distribution of SpaceX's launch sites and their proximity to coasts, which are crucial for certain mission parameters like payload capacity and safety during first-stage recovery.

SpaceX Launch Sites with Color-Coded Landing Outcomes

- SpaceX Launch Sites with Color-Coded Landing Outcomes
- **Green markers** for successful landings.
- **Red markers** for failed landings.
- **Yellow markers** for planned landings (e.g., controlled ocean landings)
- **Geographical Insight**: All launch sites are close to coastlines, enabling safe landings or planned ocean recoveries.
- **Outcome Distribution**: You can visually observe which sites have a higher success rate by noting the marker colors.

Proximity Analysis of Selected Launch Site

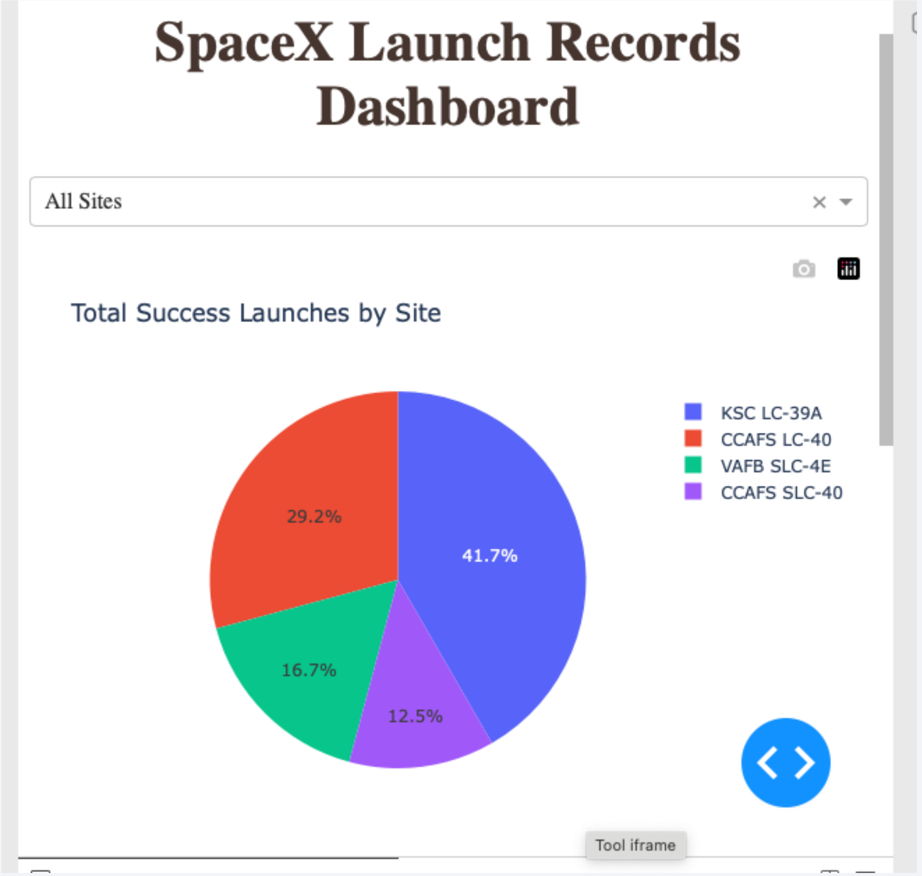


- **Proximity Analysis of Selected Launch Site**
- The map centers on a specific SpaceX launch site with a marker indicating its precise location.
- The launch site is highlighted for easy identification.
- Markers or lines representing proximities to nearby **railways**, **highways**, and **coastlines**.

Section 4

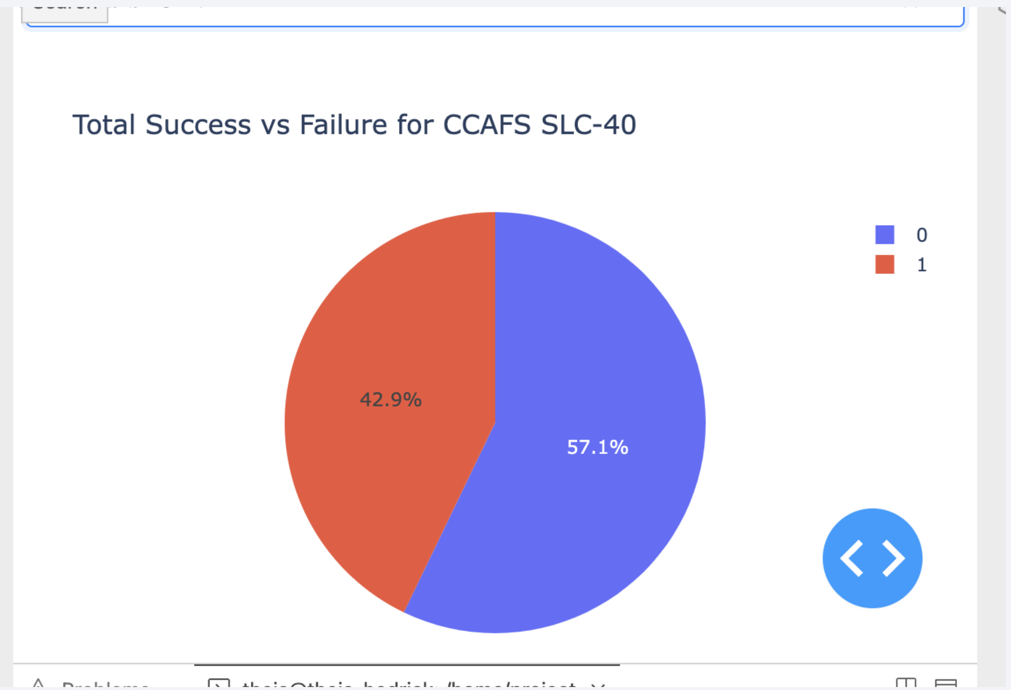
Build a Dashboard with Plotly Dash

Lunch Success Count Across all Sites



- The pie chart visually represents the total number of successful launches for each launch site.
- Each segment corresponds to a specific launch site, with its size proportional to the number of successful launches
- The dominant launch site(s) can be identified, showing which site is the most reliable or frequently used for missions.
- Sites with lower success counts may require further analysis to identify potential improvements or understand mission-specific challenges.

Pie Chart of Launch Success Ratio for Highest Performing Launch Site



- The ratio of successful launches to failures at the specified site.
- **Slices**: Each slice represents the proportion of successful and failed launches.
- **Colors**: Highlight success with a distinct color (e.g., green) to emphasize it visually.
- Highlighs the importance of understanding the success ratio for cost analysis, planning, and operational efficiency.

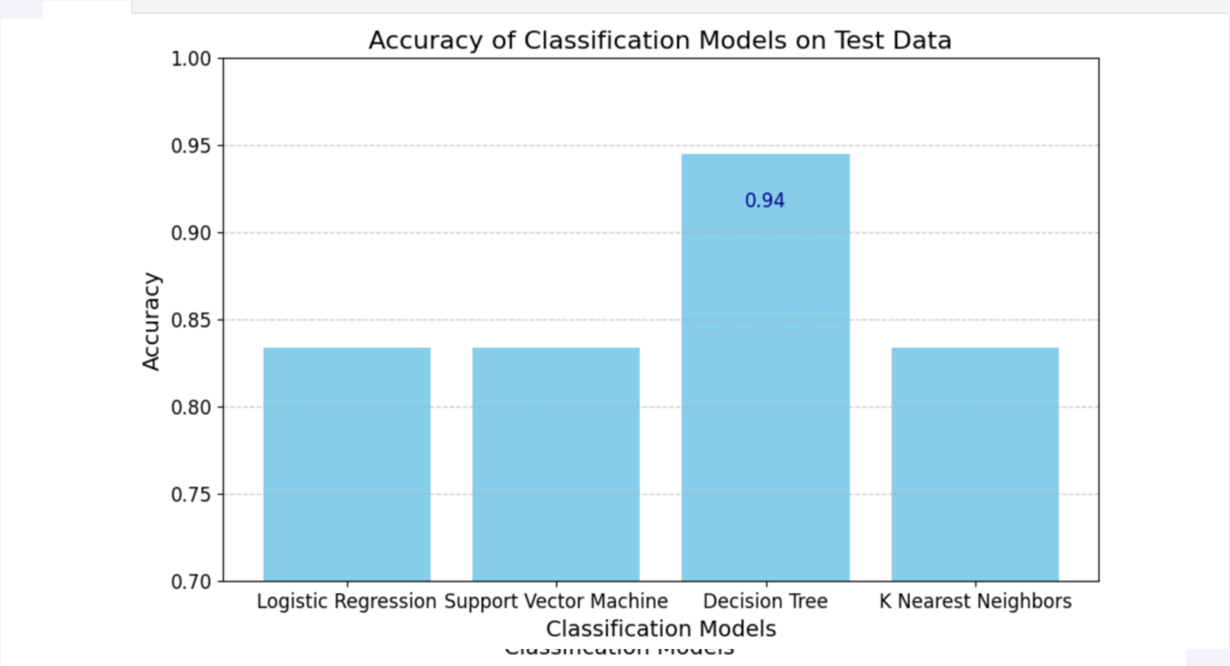
Palyload versus Launch outcome Scatter Plot for all Sites

- The scatter plot visualizes the relationship between payload mass and launch outcomes for all launch sites.
- Each point represents a launch, color-coded by the booster version.
- Payloads within a specific range (e.g., moderate weights) tend to have higher success rates.
- This scatter plot highlights crucial relationships between payload mass, booster technology, and launch outcomes, serving as a valuable tool for strategic decision-making in launch preparations.

Section 5

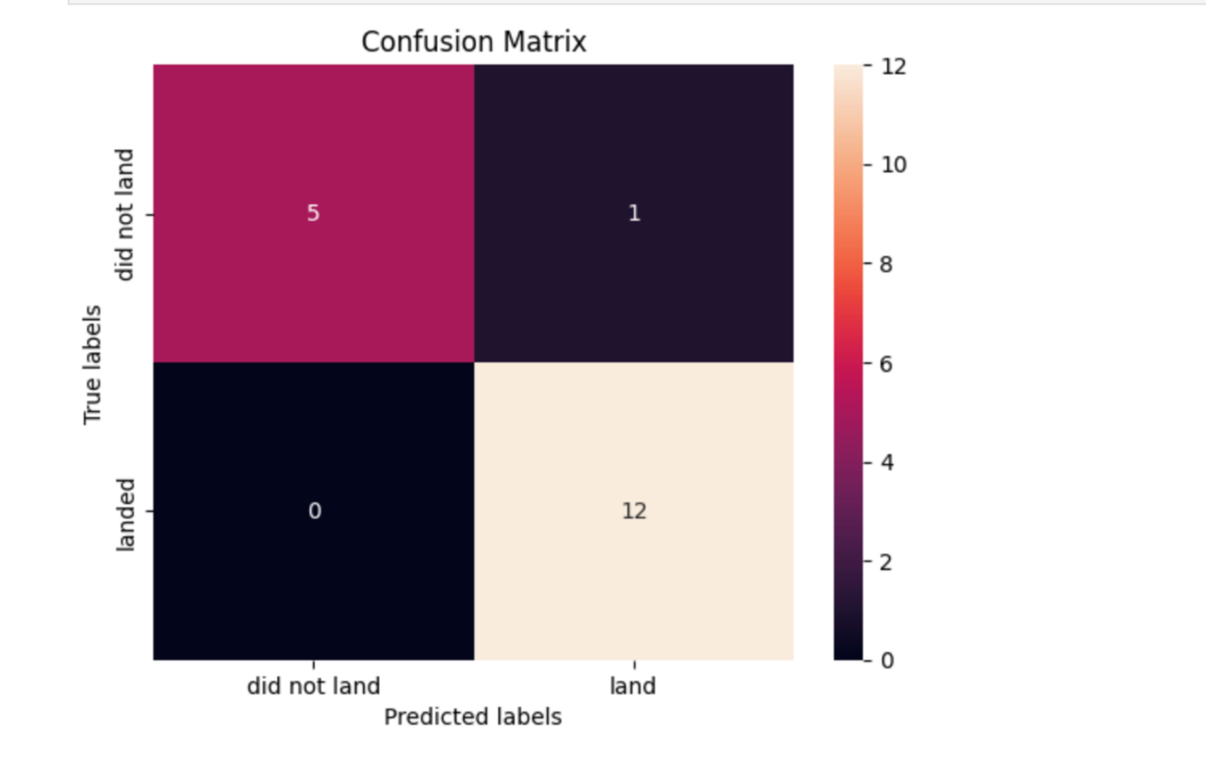
Predictive Analysis (Classification)

Classification Accuracy



- **Logistic Regression**, **Support Vector Machine**, and **K Nearest Neighbors** achieved an accuracy of **0.83**.
- **Decision Tree** achieved the highest accuracy of **0.94**.
- The chart highlights the ***Decision Tree*** model as the best-performing model with a higher classification accuracy compared to the others.

Confusion Matrix



- The confusion matrix helps evaluate the performance of the Decision Tree model

Conclusions

- The SpaceX launch data was collected from the **SpaceX REST API** and **web scraping** of Wikipedia pages using Python libraries like requests and BeautifulSoup.
- Scatter plots to analyze relationships. Line plots to track trends over time, such as yearly success rates. Bar charts to display orbit success rates and launch sites.
- SQL queries complemented EDA to extract insights like: Identifying unique launch sites and successful missions. Aggregated payload statistics for specific boosters and customers.
- Mapped all SpaceX launch sites globally to visualize geospatial distribution. Markers and overlays displayed proximity to key infrastructure and landing outcomes.
- Interactive scatter plots allowed users to dynamically explore data by payload ranges and booster versions. Pie charts and bar charts summarized launch success rates for all sites and orbit types.
- Several classification models were built, tuned, and evaluated to predict the success of SpaceX's first-stage landings: Models included **Logistic Regression**, **Support Vector Machines**, **Decision Trees**, and **K-Nearest Neighbors**.
- SpaceX has high success rates across launch sites, with **Cape Canaveral** showing the most frequent launches
- Payload size and booster versions are strong indicators of landing success, as shown by statistical and visual analysis.
- Predictive models provide SpaceX with a reliable tool to forecast mission outcomes, reducing risks and optimizing mission planning.
- The project demonstrates the value of **data-driven decision-making** for aerospace companies, especially in reducing costs and improving operational efficiency.
- The methodologies employed (API integration, data wrangling, visualization, machine learning) are highly scalable and can be extended to other industries.
- The use of interactive dashboards enhances the accessibility of insights for stakeholders, enabling real-time data exploration and decision-making.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!